Abstract

Clustering is an unsupervised learning technique which aims at grouping a set of objects into clusters so that objects in the same clusters should be similar as possible, whereas objects in one cluster should be as dissimilar as possible from objects in other clusters. Cluster analysis aims to group a collection of patterns into clusters based on similarity. A typical clustering technique uses a similarity function for comparing various data items. This paper covers the survey of various clustering techniques, the current similarity measures based on distance based clustering, explains the limitations associated with the existing clustering techniques and propose that the combination of the advantages of the existing systems can help overcome the limitations of the existing systems.

References


3. Haixun Wang, Wei Wang, Jiong Yang, Philip S. Yu, Clustering by Pattern Similarity in Large Data Sets, Proceeding SIGMOD '02 Proceedings of the 2002 ACM SIGMOD international conference on Management of data, Pages 394-405, ACM.


17. Xianchao Zhang, Xiaotong Zhang, Han Liu, Multi-Task Multi-View Clustering for Non-Negative Data, Proceedings of the Twenty-Fourth International Joint Conference on Artificial
Intelligence, IJCAI 2015.
18. Gabriella Casalino, Nicoletta Del Buono, Corrado Mencar, Subtractive clustering for
seeding non-negative matrix factorizations, Information Sciences, Volume 257, 1 February
19. Prachi Joshi, Mousami Munot, Parag Kulkarni, Madhuri Joshi, Efficient karyotyping of
metaphase chromosomes using incremental learning, IET Science, Measurement and
20. Abhishek Kumar, Hal Daume, A Co-training Approach for Multi-view Spectral
Clustering, Proceedings of the 28th International Conference on Machine Learning, Bellevue,
WA, USA, 2011.
21. Daniel John Lawson, Daniel Falush, Similarity matrices and clustering algorithms for
population identification using genetic data, Department of Mathematics, University of Bristol,
Bristol, BS8 1TW, UK, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig
Germany, March, 2012.
22. Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, Edward Y.Chang, Parallel
Spectral Clustering in Distributed Systems, Pattern Analysis and Machine Intelligence, IEEE
23. Raman Arora, Maya R. Gupta, Amol Kapila, Maryam Fazel, Similarity-based
Clustering by Left-Stochastic Matrix Factorization, Journal of Machine Learning Research 14
24. D. Kuang, C. Ding, and H. Park. Symmetric nonnegative matrix factorization for graph
26. Brian Eriksson, Gautam Dasarathy, Aarti Singh, Robert Nowak, Active Clustering:
Robust and Efficient Hierarchical Clustering using Adaptively Selected Similarities, Arxiv
27. Alina Ene, Sungjin Im, Benjamin Moseley, Fast clustering using MapReduce,
Proceeding KDD 2011 Proceedings of the 17th ACM SIGKDD international conference on
Knowledge discovery and data mining, Pages 681-689, ACM.
28. Nir Ailon, Yudong Chen, Huan Xu, Iterative and Active Graph Clustering Using Trace
Norm Minimization Without Cluster Size Constraints, Journal of Machine Learning Research 16,
2015, Pages 455-490.
29. HwanjoCheng, Yizong, Church, George M., 2000. Biclustering of expression data. In:
93-103.
30. Hwanjo Yu, Duane Searsmith, Xiaolei Li, Jiawei Han, Scalable Construction of Topic
Directory with Nonparametric Closed Termset Mining, Data Mining, 2004. ICDM ’04. Fourth
IEEE International Conference, Pages 563-566.
32. Zhengxing Huang, Zhejiang Univ., Hangzhou China, Wei Dong, Hui-long Duan, 
Haomin Li, Similarity Measure Between Patient Traces for Clinical Pathway Analysis: Problem,
Method, and Applications, Biomedical and Health Informatics, IEEE Journal of (Volume:18,
Issue: 1 ), Pages 4-14.
33. Agrawal R., Faloutsos C., Swami A. Efficient similarity search in sequence databases.
Clustering Techniques and the Similarity Measures used in Clustering: A Survey

pp. 69-84.

34. Fast Distance Metric Based Data Mining Techniques Using P-trees: k-Nearest-Neighbor classification and k-Clustering : A Thesis Submitted to the Graduate Faculty Of the North Dakota State University.


Index Terms

Computer Science

Networks

Keywords

pattern based similarity, negative data clustering, similarity measures.